UNITED STATES PATENT APPLICATION

of

R.A.L. GRIFFITHS

M.D. BARBER

P.E. JOHNSON

S.M. GILLBARD

M.D. HAYWOOD

C.D. SMITH

J. ARNOLD

T. BURKE

A.J. URQUHART

and

P. GILL

IMPROVEMENTS IN AND RELATING TO FORENSIC INDENTIFICATION

This application is a continuation of United States Patent Application Serial No. 09/706,525, filed November 3, 2000, which is a continuation of United States Patent Application Serial No. 09/498,567, filed February 4, 2000, abandoned, which is a continuation of United States Patent Application Serial No. 09/107,029, filed June 29, 1998, abandoned, which claims priority to United Kingdom Application No. 9713597.4, filed June 28, 1997, which for purposes of disclosure are incorporated herein by specific reference.

The present invention is concerned with improvements in and relating to forensic identification, particularly where based on DNA profiling.

DNA profiling offers a versatile identification technique for a wide variety of applications including, anthropological, paternity and other forensic environments. The use of such profiling is significant in determining links, or their absence, between samples. Such samples might include those taken from known individuals and/or those taken from the scene of or linked to a crime.

DNA profiling based on the use of short tandem repeats (STR) or micro satellite loci is used in such applications. STR's are a class of polymorphic markers which consist of simple tandomly repeated sequences of between 1 and 6 base pairs in length. STR's in the non-coding part of the genome are generally considered.

In the human genome STR's occur every 6 to 10 kilo bases along the DNA. The length, however, varies greatly between individuals due to mutation and provides identifying characteristics as a result.

A variety of DNA profiling systems exist, including single locus analysis and multiple locus analysis where a number of STR loci are simultaneously amplified.

In analysing the results from an unknown sample it is generally considered against a ladder marker consisting of alleles derived from actual samples. The allelic ladder provides a reference point and allows correspondence of alleles to be identified clearly.

The present invention provides new alleles and new ladders incorporating them for a variety of loci. The present invention offers an improved range and coverage of markers as a result. The ladders include a number of rare alleles offering improved identification of the alleles in an unknown sample.

According to a first aspect of the invention we provide an allelic ladder mixture comprising one or more of the following allelic ladders :-

i) an allelic ladder for locus HUMVWFA31/A comprising one or more of alleles comprising or consisting of sequences :-

TCTA TCTG TCTA $(TCTG)_4$ $(TCTA)_3$;

TCTA $(TCTG)_4$ $(TCTA)_7$; or

$(TCTA)_2$ $(TCTG)_4$ $(TCTA)_3$ TCCA $(TCTA)_3$ or at least 75% homologous thereto;

ii) an allelic ladder for locus HUMTHO1 comprising or consisting of sequence :-

$(TCAT)_4$ CAT $(TCAT)_7$ TCGT TCAT; or at least 75% homologous thereto;

iii) an allelic ladder for locus D8S1179 comprising one or more of alleles :-

$(TCTA)_8$;

$(TCTA)_2$ TCTG$(TCTA)_{16}$ or at least 75% homologous thereto;

iv) an allelic ladder for locus HUMFIBRA/FGA comprising one or more of alleles comprising or consisting of the sequences :-

$(TTTC)_3$ TTTT TTCT $(CTTT)_5$ T $(CTTT)_7$ CTCC $(TTCC)_2$;

$(TTTC)_3$ TTTT TTCT $(CTTT)_{13}$ CCTT $(CTTT)_5$ CTCC $(TTCC)_2$;

$(TTTC)_3$ TTTT TTCT $(CTTT)_{15}$ CCTT $(CTTT)_5$ CTCC $(TTCC)_2$;

$(TTTC)_4$ TTTT TT $(CTTT)_{11}$ $(CTTC)_2$ $(CTTT)_3$ CTCC $(TTCC)_4$;

$(TTTC)_4$ TTTT TT $(CTTT)_{16}$ $(CTTC)_3$ $(CTTT)_3$ CTCC $(TTCC)_4$;

$(TTTC)_4$ TTTT TT $(CTTT)_{17}$ $(CTTC)_3$ $(CTTT)_3$ CTCC $(TTCC)_4$;

$(TTTC)_4$ TTTT TT $(CTTT)_8$ $(CTGT)_4$ $(CTTT)_{13}$ $(CTTC)_4$ $(CTTT)_3$ CTCC $(TTCC)_4$;

$(TTTC)_4$ TTTT TT $(CTTT)_8$ $(CTGT)_5$ $(CTTT)_{13}$ $(CTTC)_4$ $(CTTT)_3$ CTCC $(TTCC)_4$;

$(TTTC)_4$ TTTT TT $(CTTT)_{11}$ $(CTGT)_3$ $(CTTT)_{14}$ $(CTTC)_3$ $(CTTT)_3$ CTCC $(TTCC)_4$;

$(TTTC)_4$ TTTT TT $(CTTT)_{10}$ $(CTGT)_5$ $(CTTT)_{13}$ $(CTTC)_4$ $(CTTT)_3$ CTCC $(TTCC)_4$;

$(TTTC)_4$ TTTT TT $(CTTT)_{12}$ $(CTGT)_5$ $(CTTT)_{14}$ $(CTTC)_3$ $(CTTT)_3$ CTCC $(TTCC)_4$; or

$(TTTC)_4$ TTTT TT $(CTTT)_{14}$ $(CTGT)_3$ $(CTTT)_{14}$ $(CTTC)_4$ $(CTTT)_3$ CTCC $(TTCC)_4$; or at least 75% homologous thereto;

v) an allelic ladder for locus D21S11 comprising one or more of alleles comprising or consisting of sequences :-

$(TCTA)_4$ $(TCTG)_6$ $(TCTA)_3$ TA$(TCTA)_3$ TCA $(TCTA)_2$ TCCATA $(TCTA)_6$ TCGTCT;

$(TCTA)_5$ $(TCTG)_6$ $(TCTA)_3$ TCA $(TCTA)_2$ TCCATA $(TCTA)_9$ TCGTCT;

$(TCTA)_5$ $(TCTG)_6$ $(TCTA)_3$ TCA $(TCTA)_2$ TCCATA $(TCTA)_{10}$ TCGTCT;

$(TCTA)_4$ $(TCTG)_6$ $(TCTA)_3$ TA $(TCTA)_3$ TCA $(TCTA)_2$ TCCATA $(TCTA)_8$ TCGTCT;

$(TCTA)_5$ $(TCTG)_5$ $(TCTA)_3$ TA $(TCTA)_3$ TCA $(TCTA)_2$ TCCATA $(TCTA)_9$ TCGTCT;

$(TCTA)_4$ $(TCTG)_6$ $(TCTA)_3$ TA $(TCTA)_3$ TCA $(TCTA)_2$ TCCATA $(TCTA)_{10}$ TCGTCT;

$(TCTA)_4$ $(TCTG)_6$ $(TCTA)_3$ TA $(TCTA)_3$ TCA $(TCTA)_2$ TCCATA $(TCTA)_{11}$ TCGTCT;

$(TCTA)_6$ $(TCTG)_5$ $(TCTA)_3$ TA $(TCTA)_3$ TCA $(TCTA)_2$ TCCATA $(TCTA)_{11}$ TCGTCT;

$(TCTA)_5$ $(TCTG)_6$ $(TCTA)_3$ TA $(TCTA)_3$ TCA $(TCTA)_2$ TCCATA $(TCTA)_{12}$ TCGTCT;

$(TCTA)_5$ $(TCTG)_6$ $(TCTA)_3$ TA $(TCTA)_3$ TCA $(TCTA)_2$ TCCATA $(TCTA)_{11}$ TA TCTA TCGTCT;

$(TCTA)_5$ $(TCTG)_6$ $(TCTA)_3$ TA $(TCTA)_3$ TCA $(TCTA)_2$ TCCATA $(TCTA)_{12}$ TA TCTA TCGTCT;

$(TCTA)_5$ $(TCTG)_6$ $(TCTA)_3$ TA $(TCTA)_3$ TCA $(TCTA)_2$ TCCATA $(TCTA)_{13}$ TA TCTA TCGTCT;

$(TCTA)_5$ $(TCTG)_6$ $(TCTA)_3$ TA $(TCTA)_3$ TCA $(TCTA)_2$ TCCATA $(TCTA)_{14}$ TATCTA TCGTCT;

$(TCTA)_{10}$ $(TCTG)_5$ $(TCTA)_3$ TA $(TCTA)_3$ TCA $(TCTA)_2$ TCCATA $(TCTA)_{12}$ TCGTCT;

$(TCTA)_{11}$ $(TCTG)_5$ $(TCTA)_3$ TA $(TCTA)_3$ TCA $(TCTA)_2$ TCCATA $(TCTA)_{12}$ TCGTCT;

$(TCTA)_{11}$ $(TCTG)_5$ $(TCTA)_3$ TA $(TCTA)_3$ TCA $(TCTA)_2$ TCCATA $(TCTA)_{13}$ TCGTCT; or

$(TCTA)_{13}$ $(TCTG)_5$ $(TCTA)_3$ TA $(TCTA)_3$ TCA $(TCTA)_2$ TCCATA $(TCTA)_{12}$ TCGTCT; or at least 75% homologous thereto;

vi) an allelic ladder for locus D18S51 comprising an allele comprising or consisting of sequence :-

$(AGAA)_8$; or at least 75% homologous thereto.

Preferably the mixture includes allelic ladders for a plurality of loci. It is particularly preferred that the mixture include allelic ladders for at least four loci. Preferably the mixture includes allelic ladders for a plurality of loci selected from HUMVWFA31/A, HUMTHO1, D8S1179, HUMFIBRA/FGA, D21S11 and D18S51. Preferably the mixture includes allelic ladders for at least four of these loci. In its most preferred form the mixture includes allelic ladders for all of these loci.

Preferably the mixture includes an amelogenin sex test.

Preferably one or more of the allelic ladders in the mixture includes at least 7 alleles and more preferably at least 12 alleles. Preferably a plurality, and particularly all, the allelic ladders of the mixture include at least 8 and more preferably at least 10 alleles.

Preferably one or more or all of the ladders, if present in the mixture may be provided such that: the HUMVWFA31/A allelic ladder includes at least 9, more preferably 11 and ideally 12 alleles; the HUMTHO1 allelic ladder includes at least 7, more preferably 9 and ideally 10 alleles; the D8S1179 allelic ladder

includes at least 9, more preferably 12 and ideally 13 alleles; the HUMFIBRA/FGA allelic ladder includes at least 18, more preferably 26 and ideally 28 alleles or is present as HUMFIBRA/FGA/LW and HUMFIBRA/FGA/HW with the HUMFIBRA/FGA/LW ladder including at least 16 more preferably 18 and ideally 20 alleles, the HUMFIBRA/FGA/HW ladder including at least 6, more preferably at leats 7 and ideally 8 alleles; the D21S11 allelic ladder includes at least 14, more preferably 16 and ideally 17 alleles; and the D18S51 ladder includes at least 15, more preferably 19 and ideally 20 alleles.

Preferably one or more of the allelic ladders in the mixture comprises at least 4 pairs of alleles 4 base pairs from each other. More preferably at least 10 pairs, and ideally at least 12 pairs of alleles are so provided. Preferably one or more or all the allelic ladders, if present in the mixture, may be provided such that: the HUMVWFA31/A allelic ladder includes at least 7, more preferably 10 and ideally 11 pairs of alleles 4 base pairs from each other; the HUMTHO1 allelic ladder includes at least 5, more preferably 6 and ideally 7 pairs of alleles 4 base pairs from each other; the D8S1179 allelic ladder includes at least 8, more preferably 11 and ideally 12 pairs of alleles 4 base pairs from each other; the HUMFIBRA/FGA allelic ladder includes at least 17, more preferably 20 and ideally 23 pairs of alleles 4 base pairs from each other; the D21S11 allelic ladder includes at least 3 and ideally 4 pairs of alleles 4 base pairs from each other; and the D18S51 ladder includes at least 13, more preferably 18 and ideally 19 pairs of alleles 4 base pairs from each other. The D21S11 allelic ladder may, or may further include, at least 8, more preferably 11 and ideally 12 pairs of alleles 8 base pairs from each other.

Preferably the allele sequences have at least 85% homogeneity with the listed sequences More preferable levels of even 90% or at least 95% may be provided. Ideally the exact sequences listed are included within the alleles. In their most preferred form the alleles consist of the listed sequences.

The alleles may further include flanking sequences, ie. between the primer and STR.

Preferably the HUMVWFA31/A ladder includes alleles ranging from 130, more preferably 126 and ideally 122 base pairs upwards and/or from 166 base pairs downwards. Preferably the HUMTHO1 ladder includes alleles ranging from 150 base pairs upwards and/or 189 base pairs downwards. Preferably the D8S1179 ladder includes alleles ranging from 157 base pairs upwards and/or 201, and more preferably 205 base pairs downwards. Preferably the HUMFIBRA/FGA ladder includes alleles ranging from 173 base pairs upwards and/or 298, more preferably 302 and ideally 310 base pairs downwards. Preferably the D21S11 ladder includes alleles ranging from 203 base pairs upwards and/or 255 or more preferably 259 base pairs downwards. Preferably the D18S51 ladder includes alleles ranging from 270 or more preferably 266 base pairs upwards and/or 326 or 330 or 334 or 338 or even 342 downwards.

According to a second aspect of the invention we provide an allelic ladder mixture comprising an allelic ladder for one or more of the following loci, with lowest molecular weight allele and/or uppermost molecular weight allele as follows :-

	Locus	Low MW allele	High MW allele
a)	HUMVWFA31/A	10	21
b)	HUMTH01	4	13.3
c)	D8S1179	7	19
d)	HUMFIBRA/FGA	16.1	50.2
e)	D21S11	53	81
f)	D18S51	8	27

Preferably one or more of the loci ladders have both the upper and lower limits specified. Preferably all the loci ladders have the full ranges listed.

Preferably the mixture includes allelic ladders for a plurality of loci. It is particularly preferred that the mixture include allelic ladders for at least four loci. Preferably the mixture

includes allelic ladders for a plurality of loci selected from HUMVWFA31/A, HUMTHO1, D8S1179, HUMFIBRA/FGA, D21S11 and D18S51. Preferably the mixture includes allelic ladders for at least four of these loci. In its most preferred form the mixture includes allelic ladders for all of these loci.

The intervals of alleles in the ladders and/or number of alleles in the ladders may be as specified in the first aspect of the invention. This aspect may include any of the other features specified elsewhere in the application.

The ladder mixtures of the first and/or second aspect of the invention may further include one or more of PARR buffer, primer(s), or Taq polymerase.

According to a third aspect of the invention we provide a method of analysing one or more samples comprising :-

a) obtaining genomic DNA from the sample;

b) amplifying the DNA;

c) obtaining an indication of one or more of the constituent parts of the sample; and comparing the indications with an allelic ladder mixture comprising one or more of the following allelic ladders :-

i) an allelic ladder for locus HUMVWFA31/A comprising one or more of alleles comprising or consisting of sequences :-

TCTA TCTG TCTA $(TCTG)_4$ $(TCTA)_3$;

TCTA $(TCTG)_4$ $(TCTA)_7$; or

$(TCTA)_2$ $(TCTG)_4$ $(TCTA)_3$ TCCA $(TCTA)_3$

ii) an allelic ladder for locus HUMTHO1 comprising or consisting of sequence :-

$(TCAT)_4$ CAT $(TCAT)_3$ TCGT TCAT;

iii) an allelic ladder for locus D8S1179 comprising one or more of alleles comprising or consisting of sequences :-

$(TCTA)_8$; or

$(TCTA)_2$ TCTG $(TCTA)_{16}$;

iv) an allelic ladder for locus HUMFIBRA/FGA comprising one or more of alleles comprising or consisting of the sequences :-

$(TTTC)_3$ TTTT TTCT $(CTTT)_5$ T $(CTTT)_3$ CTCC $(TTCC)_2$;

$(TTTC)_3$ TTTT TTCT $(CTTT)_{13}$ CCTT $(CTTT)_5$ CTCC $(TTCC)_2$;

$(TTTC)_3$ TTTT TTCT $(CTTT)_{16}$ CCTT $(CTTT)_5$ CTCC $(TTCC)_2$;

$(TTTC)_4$ TTTT TT $(CTTT)_{15}$ $(CTTC)_3$ $(CTTT)_3$ CTCC $(TTCC)_4$;

$(TTTC)_4$ TTTT TT $(CTTT)_{16}$ $(CTTC)_3$ $(CTTT)_3$ CTCC $(TTCC)_4$;

$(TTTC)_4$ TTTT TT $(CTTT)_{17}$ $(CTTC)_3$ $(CTTT)_3$ CTCC $(TTCC)_4$;

$(TTTC)_4$ TTTT TT $(CTTT)_8$ $(CTGT)_4$ $(CTTT)_{13}$ $(CTTC)_4$ $(CTTT)_3$ CTCC $(TTCC)_4$;

$(TTTC)_4$ TTTT TT $(CTTT)_8$ $(CTGT)_5$ $(CTTT)_{13}$ $(CTTC)_4$ $(CTTT)_3$ CTCC $(TTCC)_4$;

$(TTTC)_4$ TTTT TT $(CTTT)_{11}$ $(CTGT)_3$ $(CTTT)_{14}$ $(CTTC)_3$ $(CTTT)_3$ CTCC $(TTCC)_4$;

$(TTTC)_4$ TTTT TT $(CTTT)_{10}$ $(CTGT)_5$ $(CTTT)_{13}$ $(CTTC)_4$ $(CTTT)_3$ CTCC $(TTCC)_4$;

$(TTTC)_4$ TTTT TT $(CTTT)_{12}$ $(CTGT)_5$ $(CTTT)_{14}$ $(CTTC)_3$ $(CTTT)_3$ CTCC $(TTCC)_4$; or

$(TTTC)_4$ TTTT TT $(CTTT)_{14}$ $(CTGT)_3$ $(CTTT)_{14}$ $(CTTC)_4$ $(CTTT)_2$ CTCC $(TTCC)_4$;

v) an allelic ladder for locus D21S11 comprising one or more of alleles comprising or consisting of sequences :-

$(TCTA)_4$ $(TCTG)_6$ $(TCTA)_3$ TA$(TCTA)_3$ TCA $(TCTA)_2$ TCCATA $(TCTA)_6$ TCGTCT;

$(TCTA)_5$ $(TCTG)_6$ $(TCTA)_3$ TCA $(TCTA)_2$ TCCATA $(TCTA)_9$ TCGTCT;

$(TCTA)_5$ $(TCTG)_6$ $(TCTA)_3$ TCA $(TCTA)_2$ TCCATA $(TCTA)_{10}$ TCGTCT;

$(TCTA)_4$ $(TCTG)_6$ $(TCTA)_3$ TA $(TCTA)_3$ TCA $(TCTA)_2$ TCCATA $(TCTA)_8$ TCGTCT;

$(TCTA)_5$ $(TCTG)_5$ $(TCTA)_3$ TA $(TCTA)_3$ TCA $(TCTA)_2$ TCCATA $(TCTA)_9$ TCGTCT;

$(TCTA)_4$ $(TCTG)_6$ $(TCTA)_3$ TA $(TCTA)_3$ TCA $(TCTA)_2$ TCCATA $(TCTA)_{10}$ TCGTCT;

$(TCTA)_4$ $(TCTG)_6$ $(TCTA)_3$ TA $(TCTA)_3$ TCA $(TCTA)_2$ TCCATA $(TCTA)_{11}$ TCGTCT;

$(TCTA)_6$ $(TCTG)_5$ $(TCTA)_3$ TA $(TCTA)_3$ TCA $(TCTA)_2$ TCCATA $(TCTA)_{11}$ TCGTCT;

$(TCTA)_5$ $(TCTG)_6$ $(TCTA)_3$ TA $(TCTA)_3$ TCA $(TCTA)_2$ TCCATA $(TCTA)_{12}$ TCGTCT;

$(TCTA)_5$ $(TCTG)_6$ $(TCTA)_3$ TA $(TCTA)_3$ TCA $(TCTA)_2$ TCCATA $(TCTA)_{11}$ TA TCTA TCGTCT;

$(TCTA)_5$ $(TCTG)_6$ $(TCTA)_3$ TA $(TCTA)_3$ TCA $(TCTA)_2$ TCCATA $(TCTA)_{12}$ TA TCTA TCGTCT;

$(TCTA)_5$ $(TCTG)_6$ $(TCTA)_3$ TA $(TCTA)_3$ TCA $(TCTA)_2$ TCCATA $(TCTA)_{13}$ TA TCTA TCGTCT;

$(TCTA)_5$ $(TCTG)_6$ $(TCTA)_3$ TA $(TCTA)_3$ TCA $(TCTA)_2$ TCCATA $(TCTA)_{14}$ TATCTA TCGTCT;

$(TCTA)_{10}$ $(TCTG)_5$ $(TCTA)_3$ TA $(TCTA)_3$ TCA $(TCTA)_2$ TCCATA $(TCTA)_{12}$ TCGTCT;

$(TCTA)_{11}$ $(TCTG)_5$ $(TCTA)_3$ TA $(TCTA)_3$ TCA $(TCTA)_2$ TCCATA $(TCTA)_{12}$ TCGTCT;

$(TCTA)_{11}$ $(TCTG)_5$ $(TCTA)_3$ TA $(TCTA)_3$ TCA $(TCTA)_2$ TCCATA $(TCTA)_{13}$ TCGTCT; or

$(TCTA)_{13}$ $(TCTG)_5$ $(TCTA)_3$ TA $(TCTA)_3$ TCA $(TCTA)_2$ TCCATA $(TCTA)_{12}$ TCGTCT;

vi) an allelic ladder for locus D18S51 comprising an allele comprising or consisting of sequence :-

$(AGAA)_8$;

including allelic ladders or alleles 75% homologous thereto.

The allelic ladder mixture may possess other features specified in the first or second aspects of the invention or elsewhere in this application.

Preferably the DNA sample is one or more of a sample taken from the scene of a crime, a sample associated with the scene of a crime, a sample obtained from a suspect. a sample obtained from a human under consideration (for instance for paternity or

maternity analysis) or a reference sample. The sample may be in the form of blood, hair, skin or bodily fluid.

Preferably the sample is amplified using a polymerase chain reaction. Preferably primers for one or more of loci HUMVWFA31/A, HUMTHO1, D8S1179, HUMFIBRA/FGA, D21S11 or D18S51 are employed. The primers may be dye or otherwise labelled.

According to a fourth aspect of the invention we provide one or more alleles comprising or consisting of sequences

TCTA TCTG TCTA $(TCTG)_4$ $(TCTA)_3$;

TCTA $(TCTG)_4$ $(TCTA)_7$;

$(TCTA)_2$ $(TCTG)_4$ $(TCTA)_3$ TCCA $(TCTA)_3$;

$(TCAT)_4$ CAT $(TCAT)_7$ TCGT TCAT;

$(TCTA)_8$;

$(TCTA)_2$ TCTG $(TCTA)_{16}$;

$(TTTC)_3$ TTTT TTCT $(CTTT)_5$ T $(CTTT)_3$ CTCC $(TTCC)_2$;

$(TTTC)_3$ TTTT TTCT $(CTTT)_{13}$ CCTT $(CTTT)_5$ CTCC $(TTCC)_2$;

$(TTTC)_3$ TTTT TTCT $(CTTT)_{16}$ CCTT $(CTTT)_5$ CTCC $(TTCC)_2$;

$(TTTC)_4$ TTTT TT $(CTTT)_{15}$ $(CTTC)_3$ $(CTTT)_3$ CTCC $(TTCC)_4$;

$(TTTC)_4$ TTTT TT $(CTTT)_{16}$ $(CTTC)_3$ $(CTTT)_3$ CTCC $(TTCC)_4$;

$(TTTC)_4$ TTTT TT $(CTTT)_{17}$ $(CTTC)_3$ $(CTTT)_3$ CTCC $(TTCC)_4$;

$(TTTC)_4$ TTTT TT $(CTTT)_8$ $(CTGT)_4$ $(CTTT)_{13}$ $(CTTC)_4$ $(CTTT)_3$ CTCC $(TTCC)_4$;

$(TTTC)_4$ TTTT TT $(CTTT)_8$ $(CTGT)_5$ $(CTTT)_{13}$ $(CTTC)_4$ $(CTTT)_3$ CTCC $(TTCC)_4$;

$(TTTC)_4$ TTTT TT $(CTTT)_{11}$ $(CTGT)_3$ $(CTTT)_{14}$ $(CTTC)_3$ $(CTTT)_3$ CTCC $(TTCC)_4$;

$(TTTC)_4$ TTTT TT $(CTTT)_{10}$ $(CTGT)_5$ $(CTTT)_{13}$ $(CTTC)_4$ $(CTTT)_3$ CTCC $(TTCC)_4$;

$(TTTC)_4$ TTTT TT $(CTTT)_{12}$ $(CTGT)_5$ $(CTTT)_{14}$ $(CTTC)_3$ $(CTTT)_3$ CTCC $(TTCC)_4$;

$(TTTC)_4$ TTTT TT $(CTTT)_{14}$ $(CTGT)_3$ $(CTTT)_{14}$ $(CTTC)_4$ $(CTTT)_3$ CTCC $(TTCC)_4$;

$(TCTA)_4$ $(TCTG)_6$ $(TCTA)_3$ TA$(TCTA)_3$ TCA $(TCTA)_2$ TCCATA $(TCTA)_6$ TCGTCT;

$(TCTA)_5$ $(TCTG)_6$ $(TCTA)_3$ TCA $(TCTA)_2$ TCCATA $(TCTA)_6$ TCGTCT;

$(TCTA)_5$ $(TCTG)_6$ $(TCTA)_3$ TCA $(TCTA)_2$ TCCATA $(TCTA)_{10}$ TCGTCT;

$(TCTA)_4$ $(TCTG)_6$ $(TCTA)_3$ TA $(TCTA)_3$ TCA $(TCTA)_2$ TCCATA $(TCTA)_8$ TCGTCT;

$(TCTA)_5$ $(TCTG)_5$ $(TCTA)_3$ TA $(TCTA)_3$ TCA $(TCTA)_2$ TCCATA $(TCTA)_9$ TCGTCT;

$(TCTA)_4$ $(TCTG)_6$ $(TCTA)_3$ TA $(TCTA)_3$ TCA $(TCTA)_2$ TCCATA $(TCTA)_{10}$ TCGTCT;

$(TCTA)_4$ $(TCTG)_6$ $(TCTA)_3$ TA $(TCTA)_3$ TCA $(TCTA)_2$ TCCATA $(TCTA)_{11}$ TCGTCT;

$(TCTA)_6$ $(TCTG)_5$ $(TCTA)_3$ TA $(TCTA)_3$ TCA $(TCTA)_2$ TCCATA $(TCTA)_{11}$ TCGTCT;

$(TCTA)_5$ $(TCTG)_6$ $(TCTA)_3$ TA $(TCTA)_3$ TCA $(TCTA)_2$ TCCATA $(TCTA)_{12}$ TCGTCT;

$(TCTA)_5$ $(TCTG)_6$ $(TCTA)_3$ TA $(TCTA)_3$ TCA $(TCTA)_2$ TCCATA $(TCTA)_{11}$ TA TCTA TCGTCT;

$(TCTA)_5$ $(TCTG)_6$ $(TCTA)_3$ TA $(TCTA)_3$ TCA $(TCTA)_2$ TCCATA $(TCTA)_{12}$ TA TCTA TCGTCT;

$(TCTA)_5$ $(TCTG)_6$ $(TCTA)_3$ TA $(TCTA)_3$ TCA $(TCTA)_2$ TCCATA $(TCTA)_{13}$ TA TCTA TCGTCT;

$(TCTA)_5$ $(TCTG)_6$ $(TCTA)_3$ TA $(TCTA)_3$ TCA $(TCTA)_2$ TCCATA $(TCTA)_{14}$ TATCTA TCGTCT;

$(TCTA)_{10}$ $(TCTG)_5$ $(TCTA)_3$ TA $(TCTA)_3$ TCA $(TCTA)_2$ TCCATA $(TCTA)_{12}$ TCGTCT;

$(TCTA)_{11}$ $(TCTG)_5$ $(TCTA)_3$ TA $(TCTA)_3$ TCA $(TCTA)_2$ TCCATA $(TCTA)_{12}$ TCGTCT;

$(TCTA)_{11}$ $(TCTG)_5$ $(TCTA)_3$ TA $(TCTA)_3$ TCA $(TCTA)_2$ TCCATA $(TCTA)_{13}$ TCGTCT;

$(TCTA)_{13}$ $(TCTG)_5$ $(TCTA)_3$ TA $(TCTA)_3$ TCA $(TCTA)_2$ TCCATA $(TCTA)_{12}$ TCGTCT; or

$(AGAA)_8$; or at least 75% homologous thereto.

Preferably the alleles are provided purified from alleles other than those of HUMVWFA31/A, HUMTH01, D8S1179, HUMFIBRA/FGA, D21511, D18551 or AMG loci.

According to a fifth aspect of the invention we provide the use of an allelic ladder according to the first aspect of the

invention and/or an allele according to the fourth aspect of the invention for comparison with a DNA analysis result.

The analysis may be a DNA profile of a sample. The profile may be based on analysis of one or more loci, in particular including one or more of HUMVWFA31/A, HUMTHO1, D8S1179, HUMFIBRA/FGA, D21S11, D18S51 or AMG. The sample may be from the scene of a crime, associated with the scene of a crime or comprise a bodily fluid sample. The sample may be used to compare two or more individuals, or samples arising therefrom, for instance in paternity and/or maternity analysis.

According to a sixth aspect of the invention we provide a method of producing an allelic ladder or mixture thereof by subjecting the ladders of the first, second or fourth aspects of the invention to PCR.

The invention will now be described, by way of example only, and with reference to the accompanying figure in which :-

Figure 1 illustrates the locus, allele designation and size for an embodiment of the invention;

Figure 2a shows an electrophoretogram of the allelic ladder for Amelogenin (AMG);

Figure 2b shows an electrophoretogram of the allelic ladder for HUMVWFA31/A;

Figure 2c shows an electrophoretogram of the allelic ladder for HUMTHO1;

Figure 2d shows an electrophoretogram of the allelic ladder for D8S1179;

Figure 2e shows an electrophoretogram of the allelic ladder for HUMFIBRA, low and high molecular weights;

Figure 2f shows an electrophoretogram of the allelic ladder for D21S11;

Figure 2g shows an electrophoretogram of the allelic ladder for D18S51;

Figure 3a shows the sequence of selected alleles forming the HUMVWFA31/A ladder;

Figure 3b shows the sequence of selected alleles forming the HUMTHO1 ladder;

Figure 3c shows the sequence of selected alleles forming the D8S1179 ladder;

Figure 3d shows the sequence of selected alleles forming the HUMFIBRA ladder;

Figure 3e shows the sequence of selected alleles forming the D21S11 ladder; and

Figure 3f shows the sequence of selected alleles forming the D18S51 ladder.

An allelic ladder mixture illustrative of the present invention is provided for loci HUMTHO1, D21S11, D8S1179, HUMVWFA31/A, HUMFIBRA/FGA and amelogenin sex test. The loci nomenclature is standard, corresponding to that used in the GENEBANK database.

The ladder mixture includes a significant number of alleles for each locus so as to provide a base line for comparison across a wide range. The loci, allelic designation and base pair sizes for the mixture are shown in Figure 1. The nomenclature for the loci is discussed in Gill et al. *1996 Int. Journal Leg. Med.* ***109*** *14-22.*

The allelic ladder mixture was presented in PARR buffer (containing Tris and 1.5mM Mg ions at pH8.0) obtained from Cambio, primers obtained from Oswell and Taq polymerase from Perkin Elmer.

Electrophoretograms for the allelic ladders are shown in Figures 2a to 2g with the allelic number designations shown.

Figures 3a to 3f show the sequences for the alleles identified in Figures 2a to 2g.

The allelic ladder mixture discussed above was produced according to the following techniques. Buccal swabs and/or bloodstains were used as the sample sources. The genomic DNA was extracted using the chelex procedure described by Walsh et al. *1991 Bio. Techniques* ***1*** *91-98.*

The recovered DNA was quantified by dot hybridisation using a higher primate specific probe, as disclosed in Walsh et al. *1992 Nucleic Acids Res.* ***20*** *5061-5065.*

Each sample was then amplified according to the conditions set out below in Table 1 with unlabelled oligonucleotide primers, the sequences for which are disclosed in Urquhart et al. *1995 Bio Techniques* ***18*** *116-121* and Oldroyd et al. *1995 Electrophoresis* ***16*** *334-337.*

TABLE 1

D18	95°C for 60 seconds 60°C for 60 seconds 72°C for 60 seconds Method: 28 cycles + 72°C for 10 minutes then hold at 4°C.	D21	94°C for 30 seconds 58°C for 60 seconds 72°C for 30 seconds Method: 26 cycles + 72°C for 10 minutes then hold at 4°C.

D8 94°C for 30 seconds 60°C for 60 seconds 72°C for 60 seconds Method: 30 cycles + 72°C for 10 minutes then hold at 4°C.	TH01 and VWA 94°C for 45 seconds 60°C for 60 seconds 72°C for 60 seconds Method: 28 cycles + 72°C for 10 minutes then hold at 4°C
FGA 93°C for 60 seconds 60°C for 60 seconds 72°C for 60 seconds Method: 30 cycles + 72°C for 10 minutes then hold at 4°C.	Amelo 93°C for 30 seconds 58°C for 75 seconds 72°C for 15 seconds Method: 30 cycles + 72°C for 10 minutes then hold at 4°C.

Individual alleles were then isolated and sequence analysis was carried out according to the methods of Barber et al. *1996 Int. Journal Leg. Med.* ***108*** *180-185* and Barber and Parkin *1996 Int. Journal Leg. Med.* ***109*** *62-65*. Both DNA strands of each allele reported were sequenced and the sequences provided in Figures 3a to 3g are the consensus results for this.

The illustrations of the alleles provided in Figures 3a to 3g follow the nomenclature recommended by the DNA commission of the International Society of Forensic Haemogenetics *1994 Int. Journal Leg. Med.* ***107*** *159-160* where the complete number of tandem repeats observed are designated by the digit. The longhand version of these sequences is provided at the end of the specific description.

To prepare the ladder cocktail amplification of the alleles is necessary. This process was performed by amplifying the purified single alleles described above using a labelled primer in each case. For the locus HUMFIBRA/FGA the ladder was produced from two separate mixes, discussed in more detail below. The primers used are disclosed in Urquhart et al. *1995 Bio Techniques* ***18*** *116-121* and Oldroyd et al. *1995 Electrophoresis* ***16*** *334-337* and were employed according to the conditions set out above in Table 1.

The singleplexs produced in this way were analysed on an Applied Biosystems 377 automated sequencer to confirm the sequences. The sequences obtained from the profiling system are one base longer than those determined form the DNA sequencing technique initially discussed above. This is due to the ability of DNA polymerase from *Thermus aquaticus* to catalyse a non-template mediated addition of a deoxyribonucleotide to the 3' hydroxyl of PCR products. This is generally known as the "n+1" product and can be generated in preference to the "n" product. The results reported here, however, refer to the "n" product rather than the "n+1" product for which the labelled primer PCR conditions have been optimised to produce.

The products of the amplification process for each locus were then diluted, mixed with one another and reanalysed to produce a single ladder for each loci having even peak heights. An initial level of 1000 Arbitary Units, AU, was increased to 1000-5000AU to give greater signal strength and volume for the ladder.

The single ladders produced in this way were then mixed together to give the cocktail discussed above. The proportions of each ladder used are controlled to give balanced peak areas. The cocktail was then validated using Applied Biosystems 373A and Applied Biosystems 377 automated sequencers with Genescan and Genotyper software.

Allelic ladders according to the invention can be prepared by applying PCR amplification techniques to a pre-existing sample of the allelic ladder mixture. Alternatively the allelic ladders can be constructed from the sequence information provided herein.

The new ladders disclosed above significantly extends the range of alleles which can be identified in any DNA profiling system.

The allelic ladder mixture is used as a control sample alongside samples from known or unknown individuals which are then segregated according to size in a gel. Alleles in the sample under test can be designated by the known alleles in the control if they are within 0.5 bases of one another. Alleles falling outside this range are estimated based on their position relative to the ladder.

Using the standard nomenclature discussed above, the ladder range for each locus, defined by the extreme low molecular weight and extreme high molecular weight alleles are :-

Locus	Low MW allele	High MW allele
HUMVWFA31/A	10	21
HUMTH01	4	13.3
D8S1179	7	19
HUMFIBRA/FGA	16.1	50.2
D21S11	53	81
D18S51	8	27

The allelic ladders also enables the identification of certain rare and hence highly discriminatory alleles, in DNA profiling thus increasing the profiling systems power.

For the various locus certain alleles are of particular significance as follows :-

Locus HUMTH01

The primers used for this locus were labelled with 6-FAM. The polymorphic region of this locus is based around a tetranucleotide motif repeat, $(TCAT)_n$, where n = 4 to 13. Particular alleles provided by the present invention include 4, 9.3, 10 and 13.3. The 9.3 and 13.3 alleles were found to have a deletion of a thiamine nucleotide at either the last base of the 4th repeat unit or the first base of the 5th repeat unit. The 13.3 allele notably possesses a non-consensus tetranucleotide (TCGT) at the 13th repeat.

Locus D21S11

The primers for this locus were also labelled with 6-FAM. The allele range extends from 53 to 81 and significantly includes alleles 53, 56, 57, 79 and 81. The polymorphic region of the D21S11 alleles is relatively complex in structure and is based around the tetranucleotide TCTR, where R is A or G (following the ambiguity codes of the Nomenclature Committee of the International Union of Biochemistry), as well as containing invariant hexa-, tri- and di-nucleotides. Both allele 54 and allele 56 deviate from this general structure in that they possess a deletion of a 14 base pair $TA(TCTA)_3$ unit immediately prior to the invariant TCA tetranucleotide.

Locus D18S51

Again primers with a 6-FAM label were used. The ladder extends to 20 distinct alleles with particularly significant alleles at 8, 9, 23, 24, 25, 26 and 27. The polymorphic region is based around a simple tetranucleotide repeat motif $(AGAA)_n$, where n is 8 to 27.

Locus D8S1179

The primers used for this locus were labelled with TET. The ladder extends from alleles 7 to 19, based on 13 separate alleles. Significant alleles include 7, 15, 18 and 19. Different generalised structures were observed between the upper and lower molecular weight ends of the ladder. In the lower molecular weight area, 161 to 177 base pairs, a simple repeat region based on the tetranucleotide TCTA exists. In the higher weight region, 181 to 201 base pairs, a compound repeat region composed of the tetranucleotide TCTR was found.

Locus HUMVWFA31/A

HEX labelled primers were used for this locus. The ladder covers alleles between 10 and 21, based on 12 alleles in total. Noteworthy alleles 10, 11 and 12 are included. The polymorphic unit is generally composed of a compound repeat following the pattern $(TCTR)_n$. For the 13 and 14 alleles a non-consensus TCCA tetranucleotide at the 10th and 11th repeats was found.

Locus HUMFIBRA/FGA

This locus also employed HEX labelled primers. As mentioned above this ladder was constructed in two separate components. A low molecular weight and high molecular weight mix was used to produce the overall ladder. The low molecular weight mix ranges from allele 16.1 to 34.2 and the high molecular weight mix from allele 42.2 to 50.2.

The low MW mix includes significant alleles 16.1, 28, 30, 30.2, 31.2, 32.2, 33.2 and 34.2. The high MW mix includes noteworthy alleles 42.2, 43.2, 44.2, 45.2, 47.2, 48.2 and 50.2.

In general the HUMFIBRA/FGA alleles have a polymorphic unit based around the compound repeat YYBY, with the alleles in the upper part of the weight range being more complex in structure than those in the lower part. Within the general framework, allele 16.1 has a T nucleotide addition in the repeat region and allele 27 has a C to T transition in the 19th repeat unit

(CTTT to CCTT). The upper MW allele range includes a stutter peak which is 4 base pairs smaller than the 50.2 allele. This artifact corresponds to allele 49.2 which has not currently been determined.

Amelogenin

Primers for this locus were once again labelled with 6-FAM. The sequence data revealed an X specific product of 105 base pairs and a Y specific product of 111 base pairs.

HUMTH01 allele sequences

13.3 $(TCAT)_4$ CAT $(TCAT)_7$ TCGT TCAT

D21S11 alleles sequences

53 $(TCTA)_4$ $(TCTG)_6$ $(TCTA)_3$ TA$(TCTA)_3$ TCA $(TCTA)_2$ TCCATA $(TCTA)_6$ TCGTCT

54 $(TCTA)_5$ $(TCTG)_6$ $(TCTA)_3$ TCA $(TCTA)_2$ TCCATA $(TCTA)_9$ TCGTCT

56 $(TCTA)_5$ $(TCTG)_6$ $(TCTA)_3$ TCA $(TCTA)_2$ TCCATA $(TCTA)_{10}$ TCGTCT

57 $(TCTA)_4$ $(TCTG)_6$ $(TCTA)_3$ TA $(TCTA)_3$ TCA $(TCTA)_2$ TCCATA $(TCTA)_8$ TCGTCT

59 $(TCTA)_5$ $(TCTG)_5$ $(TCTA)_3$ TA $(TCTA)_3$ TCA $(TCTA)_2$ TCCATA $(TCTA)_9$ TCGTCT

61 $(TCTA)_4$ $(TCTG)_6$ $(TCTA)_3$ TA $(TCTA)_3$ TCA $(TCTA)_2$ TCCATA $(TCTA)_{10}$ TCGTCT

63 $(TCTA)_4$ $(TCTG)_6$ $(TCTA)_3$ TA $(TCTA)_3$ TCA $(TCTA)_2$ TCCATA $(TCTA)_{11}$ TCGTCT

65 $(TCTA)_6$ $(TCTG)_5$ $(TCTA)_3$ TA $(TCTA)_3$ TCA $(TCTA)_2$ TCCATA $(TCTA)_{11}$ TCGTCT

67 $(TCTA)_5$ $(TCTG)_6$ $(TCTA)_3$ TA $(TCTA)_3$ TCA $(TCTA)_2$ TCCATA $(TCTA)_{12}$ TCGTCT

68 $(TCTA)_5$ $(TCTG)_6$ $(TCTA)_3$ TA $(TCTA)_3$ TCA $(TCTA)_2$ TCCATA $(TCTA)_{11}$ TA TCTA TCGTCT

70 $(TCTA)_5$ $(TCTG)_6$ $(TCTA)_3$ TA $(TCTA)_3$ TCA $(TCTA)_2$ TCCATA $(TCTA)_{12}$ TA TCTA TCGTCT

72 $(TCTA)_5$ $(TCTG)_6$ $(TCTA)_3$ TA $(TCTA)_3$ TCA $(TCTA)_2$ TCCATA $(TCTA)_{13}$ TA TCTA TCGTCT

74 $(TCTA)_5$ $(TCTG)_5$ $(TCTA)_3$ TA $(TCTA)_3$ TCA $(TCTA)_2$ TCCATA $(TCTA)_{14}$ TATCTA TCGTCT

75 $(TCTA)_{10}$ $(TCTG)_5$ $(TCTA)_3$ TA $(TCTA)_3$ TCA $(TCTA)_2$ TCCATA $(TCTA)_{12}$ TCGTCT

77 $(TCTA)_{11}$ $(TCTG)_5$ $(TCTA)_3$ TA $(TCTA)_3$ TCA $(TCTA)_2$ TCCATA $(TCTA)_{12}$ TCGTCT

79 $(TCTA)_{11}$ $(TCTG)_5$ $(TCTA)_3$ TA $(TCTA)_3$ TCA $(TCTA)_2$ TCCATA $(TCTA)_{13}$ TCGTCT

81 $(TCTA)_{13}$ $(TCTG)_5$ $(TCTA)_3$ TA $(TCTA)_3$ TCA $(TCTA)_2$ TCCATA $(TCTA)_{12}$ TCGTCT

D18S51 allele sequences

8 $(AGAA)_8$

D8S1179 allele sequences

7 $(TCTA)_8$

19 $(TCTA)_2$ TCTG $(TCTA)_{16}$

HUMVWAF31/A allele sequences

10 TCTA TCTG TCTA $(TCTG)_4$ $(TCTA)_3$

12 TCTA $(TCTG)_4$ $(TCTA)_7$

13 $(TCTA)_2$ $(TCTG)_4$ $(TCTA)_3$ TCCA $(TCTA)_3$

(Note also that the allele has an atypical 3' flanking sequence. The usual sequence is TCCA TCTA T. In this allele the sequence is $(TCCA)_2T$.

HUMFIBRA(FGA) allele sequences

16.1 $(TTTC)_3$ TTTT TTCT $(CTTT)_5$ T $(CTTT)_3$ CTCC $(TTCC)_2$

27 $(TTTC)_3$ TTTT TTCT $(CTTT)_{13}$ CCTT $(CTTT)_5$ CTCC $(TTCC)_2$

30 $(TTTC)_3$ TTTT TTCT $(CTTT)_{16}$ CCTT $(CTTT)_5$ CTCC $(TTCC)_2$

31.2 $(TTTC)_4$ TTTT TT $(CTTT)_{15}$ $(CTTC)_3$ $(CTTT)_3$ CTCC $(TTCC)_4$

32.2 $(TTTC)_4$ TTTT TT $(CTTT)_{16}$ $(CTTC)_3$ $(CTTT)_3$ CTCC $(TTCC)_4$

33.2 $(TTTC)_4$ TTTT TT $(CTTT)_{17}$ $(CTTC)_3$ $(CTTT)_3$ CTCC $(TTCC)_4$

42.2 $(TTTC)_4$ TTTT TT $(CTTT)_8$ $(CTGT)_4$ $(CTTT)_{13}$ $(CTTC)_4$ $(CTTT)_3$ CTCC $(TTCC)_4$

43.2 $(TTTC)_4$ TTTT TT $(CTTT)_8$ $(CTGT)_5$ $(CTTT)_{13}$ $(CTTC)_4$ $(CTTT)_3$ CTCC $(TTCC)_4$

44.2 $(TTTC)_4$ TTTT TT $(CTTT)_{..}$ $(CTGT)_5$ $(CTTT)_{14}$ $(CTTC)_3$ $(CTTT)_3$ CTCC $(TTCC)_4$

45.2 $(TTTC)_4$ TTTT TT $(CTTT)_{16}$ $(CTGT)_5$ $(CTTT)_{13}$ $(CTTC)_4$ $(CTTT)_3$ CTCC $(TTCC)_4$

47.2 $(TTTC)_4$ TTTT TT $(CTTT)_{..}$ $(CTGT)_5$ $(CTTT)_{14}$ $(CTTC)_3$ $(CTTT)_3$ CTCC $(TTCC)_4$

48.2 $(TTTC)_4$ TTTT TT $(CTTT)_{1.}$ $(CTGT)_5$ $(CTTT)_{14}$ $(CTTC)_4$ $(CTTT)_3$ CTCC $(TTCC)_4$